IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF NEW MEXICO

UNITED STATES OF AMERICA,

Plaintiff,

v. **No. CR-14-2563 MCA**

MELVIN RUSSELL,

Defendant.

MEMORANDUM OPINION AND ORDER

THIS MATTER is before the Court on Defendant *Melvin Russell's Motion in Limine to Exclude Testimony of Tiffany Smith Concerning DNA Analysis, and/or, In the Alternative, for a Daubert Hearing Concerning the Same* [Doc. 108]. The Court has considered the parties' submissions, testimony presented at a hearing on October 31, 2017,[1] and the relevant law, and is otherwise fully informed. For the following reasons, the Court **DENIES** Defendant's *Motion*.

I. Background

Defendant Melvin Russell is charged in a criminal complaint with aggravated sexual abuse by use of force in violation of 18 U.S.C. § 2241(a). [Doc. l] The allegations leading to the charge are set out in this Court's order on *Defendant's Motion to Suppress Statements*, filed on December 6, 2016. [Doc. 103]

Defendant now moves to exclude testimony by the Government's witness, Tiffany Smith, on DNA analysis of evidence obtained from the alleged victim. [Doc. 108] Ms.

[1] The transcript of the October 31, 2017 hearing will be designated herein by "Tr."

Smith authored two different reports of DNA analysis, one in 2014 and one in 2016. [Gov't Exh. 16; Gov't Exh. 17]

The Court held a hearing on the issue on October 31, 2017. The Government presented testimony of two witnesses, Tiffany Smith and Jerrilyn Conway. Both of these witnesses are DNA analysts from the Federal Bureau of Investigation (FBI). Defendant also presented testimony of two witnesses, Dr. Charles Brenner and Dr. Michael Spence. The substance of the relevant testimony is included in the discussion below.

II. Discussion

Rule 702 imposes a special gatekeeping obligation on this Court to ensure that expert testimony is not admitted at trial unless it is both relevant and reliable. *See Kumho Tire Co., Ltd. v. Carmichael*, 526 U.S. 137, 141 (1999); *Daubert v. Merrell-Dow Pharm., Inc.*, 509 U.S. 579, 592-93 (1993). The Federal Rules of Evidence provide that:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

Fed. R. Evid. 702 (2000). The first requirement of this rule is that the expert's specialized knowledge must "assist the trier of fact." *Id.* "Rule 702 thus dictates a common-sense inquiry of whether a juror would be able to understand the evidence without specialized knowledge concerning the subject." *United States v. Muldrow*, 19 F.3d 1332, 1338 (10th Cir. 1994).

Generally, the Court's inquiry proceeds in two steps. "In determining whether expert testimony is admissible, the district court generally must first determine whether the expert is qualified 'by knowledge, skill, experience, training, or education' to render an opinion." *United States v. Nacchio*, 555 F.3d 1234, 1241 (10th Cir. 2009) (quoting Fed. R. Evid. 702). "Second, if the expert is sufficiently qualified, the court must determine whether the expert's opinion is reliable by assessing the underlying reasoning and methodology, as set forth in *Daubert*." *Id*. While the Court is not required to consider any particular set of factors or utilize a particular procedure in making such determination with respect to expert testimony, the Court must make *some* kind of determination on the record in order to demonstrate that it has performed its gatekeeping function. *See United States v. Velarde*, 214 F.3d 1204, 1209 (10th Cir. 2000).

Where, as here, the expert is offered for her scientific knowledge, the standards for measuring reliability propounded in *Daubert* apply. In *Daubert*, the Supreme Court emphasized the terms "scientific" and "knowledge" as a basis for expert testimony. *See Daubert,* 509 U.S. at 589-90 (stating "the adjective 'scientific' implies a grounding in the methods and procedures of science"; and the word "knowledge" applies to any body of known facts or to any body of ideas inferred from such facts or accepted as truths on good grounds"). The Court concluded that "in order to qualify as 'scientific knowledge,' an inference or assertion must be derived by the scientific method"; and "[p]roposed testimony must be supported by appropriate validation—*i.e.* 'good grounds,' based on what is known." *Id.* at 590. Additionally, to be helpful to the trier of fact, the proffered

testimony must have "a valid scientific connection to" a fact in issue. *Id.* at 591-92; *see* Fed. R. Evid. 702(a) (stating the "helpfulness" requirement).

Reliability under *Daubert* is determined by looking at whether the reasoning or methodology underlying the testimony is scientifically valid, and relevance is determined by whether that reasoning or methodology properly can be applied to the facts in issue. *Smith v. Sears Roebuck and Co.*, 232 Fed.Appx. 780, 781 (10th Cir. 2007). This Court's evaluation of proposed testimony is guided by several factors:

> (1) whether a theory has been or can be tested or falsified, (2) whether the theory or technique has been subject to peer review and publication, (3) whether there are known or potential rates of error with regard to specific techniques, and (4) whether the theory or approach has "general acceptance."

Bitler v. A.O. Smith Corp., 400 F.3d 1227, 1233 (10th Cir. 2004) (citations omitted). "[T]his list is neither definitive nor exhaustive and . . . a trial judge has wide discretion both in deciding how to assess an expert's reliability and in making a determination of that reliability." *Id*. "Failure to consider one, or even any, of these factors, albeit suggestive, will not be dispositive of a district court's failure to fulfill its gatekeeping role because that role depends on the underlying factual circumstances of the particular case." *Id*.

As to the final factor—general acceptance of the methodology or theory—the Supreme Court held in *Daubert* that a lack of general acceptance may reflect poorly on the reliability of a theory, but rejected the outdated notion that "general acceptance in the particular field in which it belongs" is a prerequisite to admitting expert scientific testimony. 509 U.S. at 586-89. It stated that "a rigid general acceptance requirement

would be at odds with the liberal thrust of the Federal Rules and their general approach to relaxing the traditional barriers to opinion testimony." *Id.* at 588.

Finally, the Court must be mindful of the power of expert testimony, and carefully weigh it under Fed. R. Civ. P. 403. *See id.* (permitting the exclusion of relevant evidence "if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury"). As noted by the Supreme Court: "Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 . . . exercises more control over experts than over lay witnesses." *Daubert*, 509 U.S. at 595.

A. Is the expert qualified by knowledge, skill, experience, training, or education to render an opinion?

Defendant does not challenge Ms. Smith's qualifications. [Tr. 11:25 -12:1 (Defense Counsel: "I suppose I have no issues with her background qualifications[.]"); Doc. 108] Ms. Smith testified that she is a forensic examiner in the DNA work case unit of the FBI laboratory and had been working for the FBI since 2010. [Tr. 7:13 – 8:8] She holds a bachelor's of science degree from West Virginia University in forensic and investigative sciences and a master's of science degree from West Virginia University in biology. [Tr. 8:23 – 9:1] While a graduate student, she worked for three years in a research laboratory doing research on forensic samples, in addition to teaching courses on forensic DNA as well as DNA in general. She also did an internship at a state forensic

laboratory for three months, during which she observed the DNA casework unit as well as other units. [Tr. 8:11-20]

Ms. Smith testified that she also completed an eighteen-month long training program at the FBI prior to becoming qualified as a forensic examiner. [9:4-11] While in that training program, she worked alongside qualified forensic examiners, wrote reports, performed comparisons, and reviewed data under the direct supervision of qualified examiners. She also performed laboratory testing on mock items of evidence in the laboratory. At the end of the training period, she passed a series of oral board exercises, moot court examinations, and a competency test. [Tr. 9:4-11] She testified that she had performed DNA comparisons for "well over a thousand" cases and testified in court as an expert witness in federal and state courts approximately thirty times. [Tr. 9:15-23]

The Court finds that Ms. Smith's background, education, experience, skill, and training qualify her to testify as an expert witness.

B. Are the expert's opinions reliable?

1. Autosomal and Y Chromosome DNA

"Deoxyribonucleic acid ('DNA') is material found within cells throughout the human body that contains the information necessary to make a human being." *United States v. Kootswatewa*, No. CR1508034001PCTDLR, 2016 WL 808663, at *1 (D. Ariz. Mar. 2, 2016); *see also United States v. Shirley*, No. CR13-00622-TUC-RCC, 2015 WL

631159, at *1 (D. Ariz. Feb. 13, 2015). [Tr. 13:24 -16:14] Two types of DNA analysis are relevant here: autosomal[2] short tandem repeat (STR[3]) and Y-STR[4].

> [Autosomal] DNA analysis requires a four-step process. Step one involves extraction of the DNA using chemicals to break open the cells. Next, the quantity of the DNA obtained from the sample is determined in a process called "quantitation." Third, the examiner copies the relevant areas of the DNA using the "polymerase chain reaction" ("PCR"). Finally, the DNA is separated by size to generate a DNA profile. In criminal investigations, a DNA profile from cells found on a victim can be compared with samples from suspects to determine whether there is a match.

Kootswatewa, 2016 WL 808663, at *1; *see generally* Faigman, D., 4 Mod. Sci. Evidence § 30:2 (2017-2018 Edition). [Doc. 108. Pg. 14-17; Tr. 123:24 – 124:11; 16:18 – 17:18]

Unlike autosomal DNA, Y-DNA, which is found only in biological males, is passed, largely unchanged, from father to son, and may be common among many generations sharing common male ancestry. [Doc. 108 p. 4; Tr. 16:9-14] Thus, Y-DNA is the same among fathers, sons, brothers, uncles, male cousins, and any male in the paternal lineage. [Doc. 108 p. 5; Tr. 16:9-14] Because of the nature of Y-DNA, the utility of this type of analysis is more limited than other types of DNA analysis. Because "there are large numbers of other individuals who likely share any man's Y-chromosome, the results of Y-STR analysis only allow an examiner to determine whether a crime suspect and his male paternal relatives can be excluded as contributors." *Kootswatewa*, 2016 WL 808663, at *1.

---

[2] Pertaining to analysis of "chromosomes other than the X and Y sex chromosomes." Faigman, *supra*, Appendix 30A.

[3] "Short tandem repeat (STR)" refers to the method of DNA analysis used here. *See generally id*. § 30:2.

[4] Y-STR typing "examines loci found only on the (male) Y-chromosome." *Id*. § 30:30.

Y-STR analysis is used when a sample contains both male and female DNA and allows examiners to separate out only the male DNA. Faigman, *supra*, § 30:30. In Y-STR analysis,

> [t]he first two steps are the same as regular, autosomal DNA analysis. From there, the process differs by utilizing a special kit that seeks out only the Y-chromosome, which is found only in males, instead of any of the other 22 pairs of chromosomes found in the human body. The Y-chromosome DNA profile is then cross-referenced against a database of known samples [using] . . . the "counting method."

Kootswatewa, 2016 WL 808663, at *1; Faigman, *supra*, § 30:30 ("Just as it sounds, the counting method simply counts the number of times that the haplotype has been observed in a database."). [Doc. 108] The database used here was US Y-STR, which is maintained by the University of Central Florida. [Tr. 19:17-20]

In either case, a finding that there is a match is meaningless without statistical analyses showing the significance of the match. Faigman, *supra,* at § 30:3 ("Whatever the testing method employed, the fact that two DNA samples match one another means little unless it is possible in some manner to assess the significance of that match."). [Tr. 17:8-18] The significance of a match could be expressed as a random match probability, which "answers the following question: What is the probability that a person chosen at random from a population of unrelated people will possess a DNA profile that matches the evidence profile?" Justice Ming W. Chin, et al., Forensic DNA Evidence: Science and the Law § 5:1 (April 2017 Update). Or it could be expressed as a likelihood ratio, which is "a ratio of . . . the probability of the evidence given that it originated from a specific person of interest, versus the probability of the evidence given that it originated

from a random unrelated unknown individual."[5] [Tr. 19:1-7] *See* Chin, *supra*, § 5.5 ("[The likelihood ratio] weighs the probability that the suspect is the source of the evidence profile (P1), against the probability that the source of the evidence profile is a random individual not related to the suspect (P2).").

Because the calculation of the likelihood ratio depends on a database of known samples, the selection of the proper database, or subset within a database, is essential to obtaining an accurate likelihood ratio. Faigman, *supra*, § 30:30 (stating that because "[p]opulation substructure[6] clearly exists among Y-STR haplotypes, . . . the database used to conduct the search may be significant, and should always be documented"). Because of population substructures, analysts report likelihood ratios "by the race of the individual for whom there is a match. For example, there are categories for Caucasians, African-Americans, and Native Americans." *Kootswatewa*, 2016 WL 808663, at *2. [Tr. 27:11 – 28:3]

To analyze the significance of the DNA profiles obtained from samples, the FBI lab used STRMix, a software package that performs probabilistic genotyping. [Tr. 21:9 – 22:4] "Probabilistic genotyping refers to the use of biological modeling, statistical

[5] For instance, if the likelihood ratio was 1:100, Ms. Smith testified that "we would say the DNA typing results are one hundred times more likely, if they originated from a specific . . . person of interest, than if they originated from a random unrelated unknown individual." [Tr. 19:10-14]

[6] "'Substructuring' refers to the tendency towards decreasing genetic heterogeneity and allelic independence exhibited by ethnically homogenous, non-randomly mating populations." *United States v. Morrow*, 374 F. Supp. 2d 51, 59 n.7 (D.D.C. 2005). "In other words, a substructured population may be defined as one in which the probability of a random match between two of its members is greater than the likelihood of such a match between two members of the population at large." *Id*. (internal quotation marks and citation omitted).

theory, computer algorithms, and probability distributions to calculate likelihood ratios (LRs) and/or infer genotypes for the DNA typing results of forensic samples ('forensic DNA typing results')." SWGDAM Guidelines for Validation of Probabilistic Genotyping Systems, 06/15/15, *available at* https://www.swgdam.org/publications (last visited 11/17/17.[7] [Gov't Exh. 5] STRMix was implemented at the FBI in 2015, after the 2014 report was issued. Thus, STRMix applies only to the 2016 report.

2. Reliability Issues

At issue here are the results shown in two reports by Ms. Smith. Both reports addressed DNA obtained from eleven items:

1. Sword with tape and sheath;
2. Two vaginal swabs from the alleged victim;
3. Two cervical swabs from the alleged victim;
4. Two oral swabs from the alleged victim;
5. Two bilateral nipples and areolas swabs from the alleged victim;
6. Two left gluteal fold swabs from the alleged victim;

8.[8] Buccal sample from the alleged victim;

10. Shirt from Melvin Russell's residence;
11. Buccal sample from Melvin Russell.

[Doc. 108-1; Gov't Exh. 14, 17]

[7] SWGDAM "is a group of approximately 50 scientists representing Federal, State, and Local forensic DNA laboratories in the United States and Canada." [Gov't Exh. 11]
[8] The reports do not include items numbered 7 or 9. [Doc. 108-1; Gov't Exh. 17]

Defendant makes five arguments related to the reliability of the DNA analyses in the 2014 and 2016 reports. The Court will address each in turn.

a. Low Copy Number DNA Analysis and STRMix

Defendant did not challenge the use of STRMix in his *Motion to Exclude*. [Doc. 108] However, because testimony on "low copy number" DNA testing and interpretation of such testing using STRMix was elicited at the hearing, the Court ordered additional briefing on the parties' arguments related to "analysis of low amounts of DNA and/or mixtures and the use of probabilistic genotyping software to interpret results obtained from such mixtures." [Doc. 163] Both parties submitted an additional brief on these issues. [Doc. 167; Doc. 168] No additional evidence was submitted; briefs were limited to argument only.

The issue stems from the differences in the two reports, which arise from both additional testing and the FBI's implementation in 2015 of STRMix. [Tr. 21:15 -22:4; *Compare* Doc. 108-1, pg. 2 *with* Gov't Exh. 17, pg. 3] After STRMix was implemented at the FBI, additional testing was performed on the vaginal swabs, cervical swabs, oral swabs, and left gluteal fold swabs, and data from the original testing of the other items was reanalyzed using STRMix. [Tr. 30:4 – 31:20; 35:17-25; 36:6 -37:2; 39:12-15; 40:21-23; 43:7-12] In some cases, the results in the 2016 reports differ substantially from those in the 2014 report. [*Compare* Gov't Exh. 16 *with* Gov't Exh. 17]

Ms. Smith attributed the differences in the two reports to the implementation of STRMix. She testified that in 2014 some data was not used in calculating the significance of a match because it was below the "stochastic threshold." [Tr. 22:19-23]

"Stochastic effects" such as "allele drop-in, allele drop-out, stutter, and heterozygote peak height imbalance" may increase the possibility of error in DNA testing. *United States v. McCluskey*, 954 F. Supp. 2d 1224, 1277 (D.N.M. 2013). Therefore, "[l]aboratories may set an empirically determined threshold (usually termed a 'stochastic threshold') to establish a sample quantity which puts a sample in the potential danger zone of unreliable results." *Id.* (internal quotation marks and citation omitted)[9]. Ms. Smith stated that, prior to implementation of STRMix, "if the data was below your [stochastic] threshold, it got essentially thrown out. It was not included in the statistic," but that use of STRMix "allow[ed the FBI] to use data that previously was not interpretable." [Tr. 23:2-7] She stated that data that yielded inconclusive results using the previous method might yield different results when analyzed using STRMix. [Tr. 36:1-5]

In his supplemental brief, Defendant argues that "[t]he Government has provided no showing to satisfy the *Daubert* factors as to the specific use of low amounts of DNA in combination with STRMix, the situation presented in this case." [Doc. 168, pg. 2] He maintains that, in spite of the low amounts of DNA available on certain samples[10], no

[9] According to SWGDAM, "[a]mplification of samples containing low-level DNA may be subject to stochastic effects, where two alleles at a heterozygous locus exhibit considerably different peak heights . . . or an allele fails to amplify to a detectible level (i.e., allelic drop out)." SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, 2017, pg. 12. [Gov't Exh. 6] The stochastic threshold is defined as "the peak height value below which it is reasonable to assume that allelic drop-out may have occurred within a single-source sample." *Id.*

[10] It is not clear to the Court which samples are at issue in Defendant's arguments pertaining to low amounts of DNA. In his supplemental briefing, Defendant refers to "swabs taken from the alleged victim's breasts and gluteal folds" and later refers to a swab from the sword. [Doc. 168] Dr. Spence's testimony referenced "Item 5," which is the swab from the nipples and areolas of the alleged victim. [Tr. 220:8]

special procedures were utilized to assess the DNA and the "thresholds were simply ignored on the belief that the 'Black Box' software provided by STRMix magically accounts for stochastic effects and differentiates them from true markers." [Doc. 168, pg. 4] Finally, he argues, "If a human cannot differentiate between a stochastic effect and a true allele marker, how could a human possibly devise a mathematical model to do so?" [Doc. 168, pg. 4]

Defendant presented testimony at the hearing on low copy number testing. Dr. Spence testified that

> when you get to really, really small amounts of much compromised DNA, you can start to lose information and it's random when you're going to lose some of the information, and as you get into those lower level peaks, those can be mixed up with possible high background noise, elevated background, dye blobs, electrical spikes and these other things that we talked about, stutter peaks. . . . . It's not a real common problem, but when you get into low copy number DNA you start to see more and more of those kind of issues. They crop up as you decrease the amount of DNA that's being put into the sample.

[Tr. 216:4-17] Dr. Spence described "low copy number" as "anything below 250, 200, picograms[11]," stating, "if you're starting with that amount, you're looking at low copy number typing." [Tr. 224:4-8] He warned that in circumstances where a sample has a large amount of female DNA and a small amount of male DNA, the male DNA might be difficult to analyze. [Tr. 221:17- 222:1 (stating that "[i]t's going to be a challenging sample when you have that much more female DNA" and "there are going to be challenges as to getting good reliable typing information from such a low percentage of

[11] 200 picograms = .2 nanograms. *United States v. Davis*, 602 F. Supp. 2d 658, 668 (D. Md. 2009) (stating that "[t]here are 1,000 picograms in one nanogram. Therefore, 100 picograms is the equivalent of 0.1 nanograms.").

male DNA")] He stated that here, the DNA in item 5 included approximately nine percent male (53.5 nanograms) DNA. [Tr. 222:22-23; 248:10-24] He went on, "maybe not technically that doesn't fall below [250 picograms], but you have to remember that only 9 percent of the DNA there is for males, and we really don't know, as we go past the quantification step, how many males are in that 9 percent." [Tr. 248:2-6] Thus, although the total amount of male DNA in item 5 is more than 250 picograms, and therefore not "low copy number" DNA by Dr. Spence's definition, the Court infers from Dr. Spence's testimony and Defendant's briefing that Defendant's argument is that the amount of DNA from any single male could be less than 250 picograms.

Dr. Spence pointed to a publication by Dr. Bruce Budowle, a former senior scientist at the FBI [Tr. 220:20], which was published in 2010[12] (i.e., approximately 6 years before the STRMix developmental validation study was published in 2016), stating that

> in working with low copy number techniques or low copy number quantities of DNA, and stochastic effects that you get from those, that there needs to be caution in interpretation, that there's a lack of reliability and robustness of the DNA results that you get from such low amounts of DNA, where you can see on the electropherograms that there are issues in interpretation and he is essentially very emphatically warning scientists to be careful about this and to not tarnish the gold standard by looking at and relying upon exclusively low level amounts of DNA that may give questionable data.

[12] The Budowle publication was not entered into evidence by Defendant. The Court therefore does not consider it beyond Dr. Spence's testimony.

[Tr. 228:15 – 229:3 (paraphrasing Dr. Budowle's publication)] Dr. Spence did not discuss whether Dr. Budowle's publication addresses analysis of low copy number DNA using probabilistic genotyping or STRMix.

At the hearing, Dr. Spence noted that data that had been inconclusive in 2014 was reported as conclusive in the 2016 report, due to analysis of the same data using the new methodology—STRMix. [Gov't Exh. 17, pg. 2; Tr. 224:20 – 225:2] He agreed with the conclusion in the 2014 report as it pertained to Item 5, which was that "the autosomal DNA foreign to [the alleged victim] obtained from [I]tem 5 is not suitable for matching purposes" and that "[n]o comparison information for [I]tem 5 can be provided for [Defendant]." [Doc. 108-1, pg. 2; Tr. 230:3-5] He observed that the conclusion in the 2016 report, in contrast to the 2014 report, stated that "[Defendant] is the source of the autosomal DNA unlike [the alleged victim] obtained from [I]tem 5." [Gov't Exh. 17, pg. 2; Tr. 224:20 – 225:2] He stated that he had

> never seen [a change in the analysis from inconclusive to a match] in an instance where no such reanalysis of the actual physical materials was performed, it was just a reanalysis of the data. I've never seen a jump from inconclusive to, in this case, not a conclusion that the person cannot be excluded. It's a conclusion of a match and no doubts. Because there's really no indication that there's a stats calculation that there's a low number, a low level statistic that's saying how many other people might match. This is just saying it's his DNA. And I've never seen that before, absent reanalysis and more physical work done on the actual evidence.

[Tr. 225:14-25]

However, Dr. Spence also testified that he had "never seen a case analyzed by STRmix in any previous case, and I'm not an expert on developing software for mixture analysis, . . . and I'm not an expert in that area." [Tr. 241:14-19] He went on, "I have a

lack of expertise in understanding the inner workings of STRmix or other tools like that, and . . . I can't really render an opinion on whether [STRMix is] improving the technology or making big mistakes with the technology. I have no opinion on that. I can't." [Tr. 244:5-12]

The Government presented evidence on the reliability of STRMix and its ability to analyze low copy number DNA. Ms. Smith testified that STRMix had been validated both by the developers and by the FBI in two different published, peer-reviewed studies, and that STRMix is generally considered reliable in the field. [Tr. 21:19 -22:12; 37:3-9; 38:9-12; 39:3-6] The developmental and internal validation[13] study papers, published in 2016 and 2017 respectively, were admitted into evidence. [Gov't Exh. 3 and 7] Both studies complied with the Scientific Working Group on DNA Analysis Methods (SWGDAM) Guidelines for Validation of Probabilistic Genotyping Systems, promulgated in 2015. [Gov't Exh. 5; Gov't Exh. 3; Gov't Exh. 7]

Ms. Conway testified that she was not aware of any peer-reviewed scientific publication that questioned use of STRMix generally. [Tr. 117:25 - 118:1] She also stated on cross-examination that the developmental and internal validation studies included testing of two, three, four, and five-person mixtures with "serial dilutions that

[13] Ms. Smith testified that "[d]evelopmental validation is the validation of a software package or a typing kit by the actual individuals who came up with this kit. . . . For probabilistic genotyping, it has to make sure it's calculat[ing] the correct statistics." [Tr. 37:12-20] She also stated that "[i]nternal validation is performed by the labs who want to use this software. So the FBI laboratory developed -- or excuse me -- performed an internal validation on STRmix prior to its use, to show how it behaves in the FBI laboratory's hands on the FBI laboratory's samples, and also allows us to see any limitations of the software." [Tr. 38:2-8]

would show the smaller contributor potentially dropping out at lower levels." [Tr. 137:9 – 138:2] The internal validation study states that it included "290 two, three, four, and five-person mixture profiles, prepared using DNA from thirteen contributors with varying individual template amounts (0.006 – 3.2 ng)[14] and total template amounts (0.019 – 4 ng)[15], were created in a range of contributor ratios." [Gov't Exh. 3, pg. 128] She stated that the proportion of DNA from major and minor contributors found in this case was included within the ranges studied in the internal validation study. [Tr. 138:3-22]

Ms. Conway also testified as to Defendant's "Black Box" theory, i.e., that the calculations made by STRMix cannot be validated because they are based on unknown formulas. She stated that STRMix was "using statistical modeling and biological modeling to estimate the percentage of the time we're seeing all of the DNA typing results and also how we are seeing drop-out" and that the benefit of probabilistic genotyping software, such as STRMix, is that "[i]t's able to model loss of information due to low level samples." [Tr. 73:7-12] She also testified that the developers of STRMix used "known mathematical modeling to generate their software package. And each of those methods they used have been peer reviewed and published in scientific journals." [Tr. 73:21-24] Similarly, the developmental validation study paper includes a list of the "scientific principles" on which STRMix is based and the publications in which the principles were discussed, and states that "[a]ll significant portions of the statistical algorithms and underlying scientific principles behind STRMix™ have been published in

[14] 6 picograms – 4200 picograms.
[15] 190 picograms – 4000 picograms.

peer reviewed journals." [Gov't Exh. 7, pg. 227] *See People v. Bullard-Daniel*, 42 N.Y.S.3d 714, 721 (N.Y. Co. Ct. 2016) (stating that "the mathematical models [behind STRMix] are themselves non-controversial and have been widely used in fields such as weather forecasting, computational biology, linguistics, genetics, engineering, physics, aeronautics, finance, and social sciences."). [*See also* Tr. 75:12-14 (stating that, for example, "STRmix uses [a process called Monte Carlo Markov Chain], and that has been around for a very long time. And it's used in physics. It's used in other non-forensic entities.")]

Based on the testimony of Ms. Smith and Ms. Conway, the Court concludes that the Government has met its burden to demonstrate that STRMix has been tested for the purpose relevant here, that such tests have been peer-reviewed and published in scientific journals, and that its analyses are based on calculations recognized as reliable in the field.

b. US Y-STR

The Court turns to Defendant's argument that the FBI's calculations of likelihood ratios are fundamentally flawed as applied to Native Americans and, therefore, are unreliable under *Daubert*. After obtaining Y-STR results from the nipple/areola swabs (Item 5) and the left gluteal fold (Item 6), Ms. Smith compared those results with the DNA profile obtained from Defendant. [Doc. 108-1, pg. 3] In her 2014 report, Ms. Smith described searching the US Y-STR database (release 4.0) and observing that the Y-STR profile from the nipple/areola swabs occurred in 12 of the 23,169 individuals in the database. [Doc. 108-1, pg. 3] More specifically, she reported that the likelihood ratio was 1:1754 in the African American population, 1:1976 in the Caucasian population,

1:1094 in the Hispanic population, and 1:161 in the Native American population. [Doc. 108-1, pg. 3]

As to the left gluteal fold swab, Ms. Smith reported in 2016 that she searched the US Y-STR database (release 4.1) and observed that the Y-STR profile from the left gluteal fold swab occurred in 219 of the 28,076 individuals in the database. [Gov't Exh. 17, pg. 3] She reported that the likelihood ratio for this swab was 1:85 in the African American population, 1:60 in the Caucasian population, 1:59 in the Hispanic population, and 1:51 in the Native American population. [Gov't Exh. 17, pg. 3]

Defendant argues that these Y-STR results are inherently unreliable because they are based on a database of "Native Americans" that is an agglomeration of people from an unknown number of tribes. [Doc. 108, pg. 10] The nature of the database, Defendant contends, renders Ms. Smith's conclusions so faulty as to be "junk science." [Doc. 108 p. 10] More specifically, Defendant describes three waves of migration from the Siberian Peninsula that populated the Americas and which resulted in three different, distinct Y-DNA haplogroups.[16] [Doc. 108, pg. 11] Defendant argues that, because many Native American tribes are "largely closed societies[with] limited genetic mixing with outside groups, . . . it is highly likely that [Y-]DNA could be widely shared amongst male members of one tribe, such as the Navajo, but the same haplotype [could] be nearly non-existent in other tribes, particularly those descending from a separate migration." [Doc. 108, pg. 11] *See Kootswatewa*, 2016 WL 808663, at *2 (discussing Y-STR "clusters").

[16] "A 'haplogroup' is a population sharing a common ancestor." *White v. Univ. of California*, 765 F.3d 1010, 1018 (9th Cir. 2014).

Because of this grouping of haplotypes among tribes, Defendant argues, use of a database that includes a variety of Native American tribes results in a misleading overstatement of the rarity of a match. In other words, if a certain haplotype is common among Navajos, but not among other tribes, the likelihood ratio of a match will be different if the match is compared to a database of only Navajos than if it is compared to a database of all Native Americans.

The Court notes that Defendant does not argue that the FBI's underlying methodology, i.e., identification of a match and the general formula for determining the likelihood ratio, is flawed. Instead, he argues that the methodology is flawed as applied to Native Americans due to specific features of the tribes that are not present in other populations. In support of this argument, Defendant presented testimony at the hearing by Dr. Charles Brenner. Dr. Brenner testified that use of a database that pooled all Native Americans together, as in this case, would be inappropriate because it would "manufacture diversity" and generate results that "would tend to frame the suspect by exaggerating the strength of the evidence." [Tr. 177:8 - 23] He pointed to a 2006 report by Michael Hammer and Alan J. Redd titled "Forensic Applications of Y Chromosome STRs and SNPs."[17] [Tr. 175:7 – 176:7] He testified that Hammer and Redd "made a

[17] The Hammer and Redd report was submitted to the United States Department of Justice, and is published on the National Criminal Justice Reference Service web site, but apparently was not published in a peer-reviewed publication. *See* Michael Hammer, et al., Forensic Applications of Y Chromosome STRs and SNPs, Doc. 211979, Award # 200-IJ-CX-K006 (Jan. 2006), available at https://www.ncjrs.gov/pdffiles1/nij/grants/211979.pdf. The report was not admitted into evidence at the hearing. Hence, the Court does not consider it beyond Dr. Brenner's testimony.

very strong recommendation that tribes—that Native Americans should be considered tribe by tribe." [Tr. 175:13-15] Dr. Brenner also testified that a study in Finland found that a "group of Finns in a province not very far from the capital had an amazing amount of commonalty among the men[,] that is the chance that two men would be the same was 1 in 10 in this region of an hour or two drive from Helsinki." [Tr. 175:16-24]

In response, the Government presented evidence that the US Y-STR database has been validated for Y-STR analysis and that use of the Native American database is appropriate in these circumstances. Ms. Conway testified that the US Y-STR database had been validated in a study that was peer-reviewed and published in 2010. [Tr. 121:12-16; Gov't Exh. 14] The paper was admitted as an exhibit. [Gov't Exh. 14] Ms. Smith testified that the use of the US Y-STR database was recommended by SWGDAM. [Tr. 20:3-5; Gov't Exh. 11] The SWGDAM guidelines for Y-STR interpretation were admitted as an exhibit. [Gov't Exh. 11] In addition, Ms. Conway testified that the FBI's calculations of likelihood ratios include an adjustment for population substructures using Θ (theta), which "describes the chance of haplotypes being the same within populations relative to the chance of them being the same between populations." [Tr. 126:21 -127:1; 147:1-18; Gov't Exh. 11, pg. 18] These calculations are recommended by SWGDAM in the Y-STR interpretation guidelines, and are based on methods described by Weir and Cockerham in 1984. [Gov't Exh. 11, pg. 13] Finally, use of Θ to adjust for population substructures has been found reliable by other courts in other contexts. *See, e.g.*, *Young v. United States*, 63 A.3d 1033, 1052 (D.C. 2013) (stating that use of the "product rule"

and a Θ correction "has been tested empirically and its scientific foundations and basic accuracy are generally accepted in the scientific community" (footnote omitted)).

To the extent Defendant relies on *Kootswatewa*, in which the court found that the use of a Native American subset of a database was inappropriate where the defendant was a member of the Hopi nation, the Court finds that *Kootswatewa* is distinguishable on a number of key facts. First, the database in that case was promulgated by Applied Biosystems and included only 105 self-identified Native Americans. *Kootswatewa*, 2016 WL 808663, at *3. The analyst "was unaware of any peer-reviewed articles examining the database or endorsing the pooling of Native American populations for purposes of Y-STR DNA analysis" and the Government did not "produce other evidence that pooling Native Americans results in a scientifically accepted category of genetic classification for purposes of Y-STR DNA analysis." *Id*. Hence, the Government did not meet its burden to demonstrate that the database and its use in that case was reliable. *Id*. at *4 (stating that "[t]he Government, which bears the burden of establishing the admissibility of its expert opinion evidence, ha[d] not shown that the data upon which [its] calculation [wa]s based [wa]s reliable"). Here, Ms. Smith testified that the database included 3300 Native American profiles. [Tr. 54:19-20] Moreover, as just discussed, both she and Ms. Conway testified that the US Y-STR database had been validated in a peer-reviewed, published study, as well as being recommended by SWGDAM. Finally, Ms. Conway testified that the calculations of the likelihood ratio accounted for population substructure within the database. [Tr. 126:21 -127:1; 147:1-18] These facts distinguish the present matter from *Kootswatewa*.

The Court finds that the Government has met its burden to demonstrate that the US Y-STR database and comparison of a match against the database of all Native American profiles in the database has been tested and validated in a scientific, peer-reviewed, published study and is considered reliable amongst experts in the field, and hence is reliable under *Daubert*. *See Bitler*, 400 F.3d at 1233.

c. 95% Confidence Interval

Both the 2014 and 2016 reports indicate in a footnote that "match probabilities are derived from the frequencies of the profile in the database after applying a 95% upper confidence limit and incorporating the population structure parameter Θ." [Doc. 108-1, pg. 4; Gov't Exh. 17, pg. 5] Defendant argues that the Government has failed to demonstrate that calculation of the 95% confidence interval is reliable and valid. [Doc. 108, pg. 7] A confidence interval is "[a]n estimate, expressed as a range, for a parameter. For estimates such as averages or rates computed from large samples, a 95% confidence interval is the range from about two standard errors below to two standard errors above the estimate. Intervals obtained this way cover the true value about 95% of the time." Comm. on Sci., Tech. & Law Policy & Global Affairs, Federal Judicial Ctr., Reference Manual on Scientific Evidence, pg. 284 (3d ed. 2011); *United States v. Shea*, 957 F. Supp. 331, 343 n.40 (D.N.H. 1997) ("Confidence intervals qualify a conclusion in an effort to account for the effect of random error by describing a range of possible results that is expected to contain the true result a given percentage of the time."). Ms. Conway testified that the FBI calculates the confidence interval using the Clopper and Pearson formula, which was developed in 1934. [Tr. 126:14-20] She also testified that use and

calculation of the interval is done according to the SWGDAM Interpretation Guidelines for Y-Chromosome STR Typing by Forensic DNA Laboratories, published in 2014. [Gov't Exh. 11; Tr. 127:10-19] Those guidelines specify use of the Clopper and Pearson formula. [Gov't Exh 11, pg. 12] Defendant did not elicit any testimony from Drs. Brenner or Spence on the confidence interval calculations.

The Court concludes that the Government sufficiently demonstrated that the formula used to calculate the confidence interval is reliable for the purpose used here.

d. The Meaning of the Likelihood Ratio

Defendant next argues that Ms. Smith fundamentally misunderstands the nature of Y-STR analysis based on the statement in the reports that "the likelihood ratio of the match probability describes . . . how much more likely the DNA match is to occur if [Defendant] is the contributor as opposed to a randomly selected individual from the same population." [Doc. 108-1 p. 2; Doc. 108-1, pg. 2] Defendant argues that what the likelihood ratio actually shows is the likelihood that Defendant *or a male relative* from a common male ancestor is the contributor. [Doc. 108 p. 8] However, the paragraph quoted by Defendant is accompanied by a footnote, which states, "Barring mutation, any male relative within the same paternal lineage has the same Y-STR profile and would also be expected to be included as a potential contributor." [Doc. 108-1, pg. 4; *see also* Gov't Exh. 17 (2016 report, including the same footnote)] Moreover, Ms. Smith clarified at the hearing that the likelihood ratio in her report of 1:161 among Native Americans meant that "the DNA typing results are at least 161 times more likely to have originated from [Defendant] *or a paternal relative to [Defendant]* than if they originated from an

unrelated unknown individual in that same population." [Tr. 27:24 - 28:3 (Emphasis added.)]

The Court concludes that Ms. Smith's testimony is not inadmissible on this basis. Rather, the extent to which the bodies of the 2014 and 2016 reports misstate the meaning of the likelihood ratio by implying that it does not include all paternal relatives is fodder for cross-examination at trial.

e. Margin of Error

In the 2014 report, a footnote states that

> [t]he uncertainty associated with a random match probability calculated using any of the FBI databases has been empirically demonstrated to be less than 10-fold in either direction (e.g., for a random match probability estimate of 1 in 10 million, the true frequency of that profile can be confidently expected to be between 1 in 1 million and 1 in 100 million).

[Doc. 108-1, pg. 4] Defendant argues that, applying this margin of error, the likelihood ratio of 1:161 could be 1:16 or 1:1600 and, therefore, the likelihood ratio is facially unreliable. [Doc. 108, pg. 8-9] However, the footnote referenced applies only to Ms. Smith's analysis of the DNA obtained from a shirt, not to the swabs of the nipples/areolas, to which the "1 in 161" likelihood ratio pertains. [Doc. 108-1, pg. 2] Moreover, a "random match probability" and a "likelihood ratio" are two different calculations. *See* SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, 2017, pg. 43 (discussing the three approaches to statistical analysis—random match probability, likelihood ratio, and combined probability of inclusion/exclusion); Chin, *supra*, § 5:5 (stating that "an alternative to the random match probability statistic is the likelihood ratio"). [Gov't Exh. 6] Thus, the

margin of error discussed in the footnote does not apply to the 1:161 likelihood ratio at all.

C. Defendant's Rule 403 Argument

Under Federal Rule of Evidence 403, relevant evidence may be excluded "if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury." Defendant argues that Ms. Smith's testimony has no probative value because it is based on inherently unreliable calculations, and hence would be misleading and unfairly prejudicial. [Doc. 108 p. 13-14] As the Court has determined that, to the contrary, the calculations are based on reliable methodologies, the Court further finds that the proposed testimony is not more prejudicial than probative. Defendant's arguments under Rule 403 are therefore unavailing.

III. Conclusion

For the foregoing reasons, the Court **DENIES** *Defendant Melvin Russell's Motion in Limine to Exclude Testimony of Tiffany Smith Concerning DNA Analysis, and/or, In the Alternative, for a Daubert Hearing Concerning the Same* [Doc. 108].

SO ORDERED this 10th day of January, 2018.

M. CHRISTINA ARMIJO
Chief United States District Judge